Carlo Nipoti, Dipartimento di Fisica e Astronomia, Università di Bologna

7/5/2014

4. Perturbations

4.1 Introduction to perturbation theory

[R05 7.1]

→ The point-mass two-body problem can be solved analytically, but when we have more then two bodies or if the point mass approximation is not justified (for instance, at least one of the two bodies is not spherical), we do not have analytic solutions.

→ In general the motion is governed by a potential function $U = U_0 + \mathcal{R}$, where U_0 is the point-mass two-body gravitational potential and $\mathcal{R}$ is the so-called disturbing function, which accounts for the presence of other bodies or for deviations from spherical symmetry in the mass distributions of the bodies.

→ Perturbation theory is a way to account for the presence of $\mathcal{R}$. We distinguish two kinds of perturbation methods: general perturbations and special perturbations.

→ In many applications the effect of $\mathcal{R}$ is at least an order of magnitude smaller than that of U_0, so both general perturbations and special perturbations can be used.

→ When the effects of $\mathcal{R}$ are comparable to those of U_0 it is not possible to use general perturbations, and special perturbation methods must be used.

→ *General perturbations:* this method exploits the fact that the orbit due to U_0 changes only slowly due to the effect of $\mathcal{R}$. So, at a given time the orbit is characterized by the osculating elements, which define the osculating ellipse (i.e. the "instantaneous" orbit due to U_0, which we assume here is an ellipse). Then equations for the variation of the elements with time are obtained and studied with analytic methods.

→ *Special perturbations (i.e. numerical integration of orbits):* given the masses of the bodies, starting from positions and velocities at a given time, positions and velocities at later times are obtained by numerical integration of the full equations of motion or of the perturbation equations (i.e. the equations for the variation of the elements).

→ General perturbation method is applicable only when the perturbation is small, but it allows to individuate the dominant perturbing terms and better understand the physical evolution. For instance, general perturbations can enable the sources of observed perturbations to be discovered, because the sources of the perturbations appear explicitly in the equations.

→ Special perturbation method is applicable to any system and over long timescales, but no attempt to isolate different perturbing terms. Fundamental tool, for instance, in studying the long-term evolution of planetary systems.

4.2 General perturbations

[R05 7.1]

→ Initial conditions: at time t_0 the osculating elements are a_0, e_0, i_0, Ω_0, ω_0 and τ_0. If $\mathcal{R} = 0$ (i.e. no perturbation) these elements are constant.

→ Due to $\mathcal{R} \neq 0$ the elements evolve and at a later time t_1 they will be a_1, e_1, i_1, Ω_1, ω_1 and τ_1.

→ The quantities $\Delta a = a_1 - a_0$ etc. are the perturbations in the time interval $\Delta t = t_1 - t_0$.

4.2.1 Lagrange's planetary equations

[R05 7.10]

Hamiltonian formulation

→ The equations that describe the evolution of the osculating elements are called Lagrange's planetary equations.

→ Here we derive Lagrange's planetary equations in the context of the Hamiltonian formulation of mechanics.

→ From the above derivation of the disturbing function (and from the equations of motion $\ddot{\mathbf{r}}_i = \nabla_i U_{0,i} + \nabla_i R_i$) we can infer the form of the corresponding Hamiltonian. Let's consider here the mass-normalized Hamiltonian $\tilde{\mathcal{H}}$ (see discussion in 2.4.2).

→ The Hamiltonian is

$$\tilde{\mathcal{H}} = \tilde{\mathcal{H}}_0 + \tilde{\mathcal{H}}_1,$$

where (dropping index i and using $\Phi_0 = -U_0$ as unperturbed gravitational potential)

$$\tilde{\mathcal{H}}_0 = \frac{1}{2}\left(\tilde{p}_x^2 + \tilde{p}_y^2 + \tilde{p}_z^2\right) + \Phi_0 = \frac{1}{2}\left(\tilde{p}_x^2 + \tilde{p}_y^2 + \tilde{p}_z^2\right) - U_0,$$

$$\tilde{\mathcal{H}}_1 = -\mathcal{R}$$

→ With this definition the canonic equations give the equations of motion derived above

$$\dot{x} = \frac{\partial \tilde{\mathcal{H}}}{\partial \tilde{p}_x} = \tilde{p}_x$$

$$\dot{\tilde{p}}_x = -\frac{\partial \tilde{\mathcal{H}}}{\partial x} \qquad \Longrightarrow \qquad \ddot{x} = \frac{\partial U_0}{\partial x} + \frac{\partial \mathcal{R}}{\partial x}$$

(here $\tilde{p}_x = \dot{x}$, because we have mass-normalized the Hamiltonian), and similarly for y and z.

Canonical coordinates

→ When considering the two-body problem (m_i and m) in Hamiltonian dynamics we have seen that it is possible to write the solution in terms of 6 canonic variables $\tilde{\alpha}_i$, $\tilde{\beta}_i$ with $i = 1, 2, 3$ such that they are all constant.

→ These constant (mass-normalized) canonical coordinates are

$$\tilde{\alpha}_1 = \tilde{E} = -\frac{GM}{2a}, \qquad \tilde{\beta}_1 = -\tau$$
$$\tilde{\alpha}_2 = \tilde{L} = \sqrt{GMa(1-e^2)}, \qquad \tilde{\beta}_2 = \omega$$
$$\tilde{\alpha}_3 = \tilde{L}_z = \sqrt{GMa(1-e^2)}\cos i, \qquad \tilde{\beta}_3 = \Omega,$$

where $M = m + m_i$, or, using $\mu = GM$,

$$\tilde{\alpha}_1 = -\frac{\mu}{2a}, \qquad \tilde{\beta}_1 = -\tau,$$
$$\tilde{\alpha}_2 = \sqrt{\mu a(1-e^2)}, \qquad \tilde{\beta}_2 = \omega,$$
$$\tilde{\alpha}_3 = \sqrt{\mu a(1-e^2)}\cos i, \qquad \tilde{\beta}_3 = \Omega.$$

→ The corresponding mass-normalized two-body Hamiltonian is null: $\tilde{\mathcal{H}}_0 = 0$. Clearly, the two-body (unperturbed) Hamiltonian $\tilde{\mathcal{H}}_0$ does not depend on $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ ($\tilde{\alpha}_i$ are the momenta and $\tilde{\beta}_i$ are the coordinates), so

$$\dot{\tilde{\alpha}}_i = -\frac{\partial \tilde{\mathcal{H}}_0}{\partial \tilde{\beta}_i} = 0,$$
$$\dot{\tilde{\beta}}_i = \frac{\partial \tilde{\mathcal{H}}_0}{\partial \tilde{\alpha}_i} = 0.$$

→ Now, in our case the Hamiltonian is in the form $\tilde{\mathcal{H}} = \tilde{\mathcal{H}}_0 + \tilde{\mathcal{H}}_1 = \tilde{\mathcal{H}}_0 - \mathcal{R}$, so the canonic variables $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ are not constant, but they vary as

$$\dot{\tilde{\alpha}}_i = -\frac{\partial \tilde{\mathcal{H}}_1}{\partial \tilde{\beta}_i} = \frac{\partial \mathcal{R}}{\partial \tilde{\beta}_i},$$
$$\dot{\tilde{\beta}}_i = \frac{\partial \tilde{\mathcal{H}}_1}{\partial \tilde{\alpha}_i} = -\frac{\partial \mathcal{R}}{\partial \tilde{\alpha}_i}.$$

→ Now we combine the above Hamilton equations to obtain the perturbation equations, i.e. Lagrange planetary equations, i.e. the equations that describe the time-variation of the orbital elements of the osculating ellipse, due to the presence of the disturbing function $\mathcal{R}$.

Hamilton's equations in terms of orbital elements

→ Hamilton equation (I)

$$\dot{\tilde{\alpha}}_1 = \frac{\partial \mathcal{R}}{\partial \tilde{\beta}_1}$$

→ So we have

$$-\frac{\partial \mathcal{R}}{\partial \tau} = \frac{\partial \mathcal{R}}{\partial \tilde{\beta}_1} = \dot{\tilde{\alpha}}_1 = \frac{\mathrm{d}}{\mathrm{d}t}\left(-\frac{\mu}{2a}\right) = \frac{\mu \dot{a}}{2a^2} = \frac{n^2 a \dot{a}}{2}$$

where we have used $\mu = n^2 a^3$ [we recall $\mu = GM = G(m + m_i)$], so

$$\dot{a} = -\frac{2}{n^2 a}\frac{\partial \mathcal{R}}{\partial \tau} \qquad (I)$$

→ Note that n appears in the expression of Lagrange's planetary equations. However, n is not an independent orbital element, and must be considered just a function of a: $n = n(a) = \sqrt{\mu/a^3}$. So $\mathrm{d}n/\mathrm{d}a = -(3/2)n/a$. Other useful relations: $\sqrt{\mu a} = na^2$ and $\sqrt{\mu/a} = na$.

→ Hamilton equation (II)

$$\dot{\tilde{\alpha}}_2 = \frac{\partial \mathcal{R}}{\partial \tilde{\beta}_2},$$

so

$$\frac{\partial \mathcal{R}}{\partial \omega} = \frac{\partial \mathcal{R}}{\partial \tilde{\beta}_2} = \dot{\tilde{\alpha}}_2 = \frac{\mathrm{d}}{\mathrm{d}t}\left[\sqrt{\mu a(1-e^2)}\right] = \frac{na\sqrt{1-e^2}}{2}\left[\dot{a} - \frac{2ae}{1-e^2}\dot{e}\right],$$

which can be written as

$$\dot{a} - \frac{2ae}{1-e^2}\dot{e} = \frac{2}{na\sqrt{1-e^2}}\frac{\partial \mathcal{R}}{\partial \omega} \qquad (II)$$

→ Hamilton equation (III)

$$\dot{\tilde{\alpha}}_3 = \frac{\partial \mathcal{R}}{\partial \tilde{\beta}_3},$$

so

$$\frac{\partial \mathcal{R}}{\partial \Omega} = \frac{\partial \mathcal{R}}{\partial \tilde{\beta}_3} = \dot{\tilde{\alpha}}_3 = \frac{\mathrm{d}}{\mathrm{d}t}\left[\sqrt{\mu a(1-e^2)}\cos i\right] = \frac{na\sqrt{1-e^2}}{2}\left[\dot{a} - \frac{2ae}{1-e^2}\dot{e}\right]\cos i - na^2\sqrt{1-e^2}\cos i \tan i \frac{\mathrm{d}i}{\mathrm{d}t} =$$

$$= \frac{na\sqrt{1-e^2}\cos i}{2}\left[\dot{a} - \frac{2ae}{(1-e^2)}\dot{e} - 2a\tan i\frac{\mathrm{d}i}{\mathrm{d}t}\right]$$

so

$$\dot{a} - \frac{2ae}{1-e^2}\dot{e} - 2a\tan i\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{2}{na\sqrt{1-e^2}\cos i}\frac{\partial \mathcal{R}}{\partial \Omega} \qquad (III)$$

→ Now we use combinations of the other three Hamilton equations

$$\dot{\tilde{\beta}}_1 = -\frac{\partial \mathcal{R}}{\partial \tilde{\alpha}_1},$$

a

$$\dot{\tilde{\beta}}_2 = -\frac{\partial \mathcal{R}}{\partial \tilde{\alpha}_2},$$

$$\dot{\tilde{\beta}}_3 = -\frac{\partial \mathcal{R}}{\partial \tilde{\alpha}_3}.$$

→ Equation (IV) for $\partial\mathcal{R}/\partial i$ (i appears only in α_3):

$$\frac{\partial\mathcal{R}}{\partial i} = \frac{\partial\mathcal{R}}{\partial\tilde{\alpha}_3}\frac{\partial\tilde{\alpha}_3}{\partial i} = -\dot{\tilde{\beta}}_3\left(-na^2\sqrt{1-e^2}\sin i\right) = \dot{\Omega}na^2\sqrt{1-e^2}\sin i$$

i.e.

$$\dot{\Omega} = \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial\mathcal{R}}{\partial i} \qquad (IV)$$

→ Equation (V) for $\partial\mathcal{R}/\partial e$ (e appears only in $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$):

$$\frac{\partial\mathcal{R}}{\partial e} = \frac{\partial\mathcal{R}}{\partial\tilde{\alpha}_2}\frac{\partial\tilde{\alpha}_2}{\partial e} + \frac{\partial\mathcal{R}}{\partial\tilde{\alpha}_3}\frac{\partial\tilde{\alpha}_3}{\partial e} = -\dot{\tilde{\beta}}_2\frac{\partial\tilde{\alpha}_2}{\partial e} - \dot{\tilde{\beta}}_3\frac{\partial\tilde{\alpha}_3}{\partial e} = -\dot{\omega}\left(-\frac{ena^2}{(1-e^2)^{1/2}}\right) - \dot{\Omega}\left(-\frac{ena^2}{(1-e^2)^{1/2}}\cos i\right) =$$

$$= \frac{ena^2}{\sqrt{1-e^2}}\left(\dot{\omega} + \dot{\Omega}\cos i\right)$$

so

$$\dot{\omega} + \dot{\Omega}\cos i = \frac{(1-e^2)^{1/2}}{ena^2}\frac{\partial\mathcal{R}}{\partial e} \qquad (V)$$

→ Equation (VI) for $\partial\mathcal{R}/\partial a$ (a appears in $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$):

$$\frac{\partial\mathcal{R}}{\partial a} = \frac{\partial\mathcal{R}}{\partial\tilde{\alpha}_1}\frac{\partial\tilde{\alpha}_1}{\partial a} + \frac{\partial\mathcal{R}}{\partial\tilde{\alpha}_2}\frac{\partial\tilde{\alpha}_2}{\partial a} + \frac{\partial\mathcal{R}}{\partial\tilde{\alpha}_3}\frac{\partial\tilde{\alpha}_3}{\partial a} = -\dot{\tilde{\beta}}_1\frac{\partial\tilde{\alpha}_1}{\partial a} - \dot{\tilde{\beta}}_2\frac{\partial\tilde{\alpha}_2}{\partial a} - \dot{\tilde{\beta}}_3\frac{\partial\tilde{\alpha}_3}{\partial a} =$$

$$= \frac{n^2a}{2}\dot{\tau} - \sqrt{1-e^2}\frac{na}{2}(\dot{\omega} + \dot{\Omega}\cos i)$$

so

$$\dot{\tau} - \frac{\sqrt{1-e^2}}{n}(\dot{\omega} + \dot{\Omega}\cos i) = \frac{2}{n^2a}\frac{\partial\mathcal{R}}{\partial a} \qquad (VI)$$

Equations for the variation of the elements

→ We now combine equations (I-VI) to obtain Lagrange's planetary equations in the form $\mathrm{d}a/\mathrm{d}t = \ldots$ etc.

(I):

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -\frac{2}{n^2a}\frac{\partial\mathcal{R}}{\partial\tau} \qquad (1)$$

(II+I):

$$\frac{\mathrm{d}e}{\mathrm{d}t} = \frac{1-e^2}{2ae}\dot{a} - \frac{1-e^2}{2ae}\frac{2}{na\sqrt{1-e^2}}\frac{\partial\mathcal{R}}{\partial\omega}$$

$$\frac{\mathrm{d}e}{\mathrm{d}t} = -\frac{\sqrt{1-e^2}}{a^2en}\left[\frac{\sqrt{1-e^2}}{n}\frac{\partial\mathcal{R}}{\partial\tau} + \frac{\partial\mathcal{R}}{\partial\omega}\right] \qquad (2)$$

(III+II):

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{1}{2a\tan i}\dot{a} - \frac{2ae}{2a\tan i(1-e^2)}\dot{e} - \frac{1}{a^2n\sin i\sqrt{1-e^2}}\frac{\partial\mathcal{R}}{\partial\Omega}$$

$$= \frac{1}{2a\tan i}\left[\dot{a} - \frac{2ae}{1-e^2}\dot{e}\right] - \frac{1}{a^2n\sin i\sqrt{1-e^2}}\frac{\partial\mathcal{R}}{\partial\Omega}$$

$$= \frac{1}{2a\tan i}\left[\frac{2}{na\sqrt{1-e^2}}\frac{\partial\mathcal{R}}{\partial\omega}\right] - \frac{1}{a^2n\sin i\sqrt{1-e^2}}\frac{\partial\mathcal{R}}{\partial\Omega}$$

$$= \frac{1}{a^2 n\sqrt{1-e^2}}\left[\frac{1}{\tan i}\frac{\partial\mathcal{R}}{\partial\omega} - \frac{1}{\sin i}\frac{\partial\mathcal{R}}{\partial\Omega}\right] \qquad (3)$$

(IV):

$$\frac{\mathrm{d}\Omega}{\mathrm{d}t} = \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial\mathcal{R}}{\partial i} \qquad (4)$$

(V+IV):

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = -\dot{\Omega}\cos i + \frac{(1-e^2)^{1/2}}{ena^2}\frac{\partial\mathcal{R}}{\partial e} = -\frac{\cos i}{na^2\sqrt{1-e^2}\sin i}\frac{\partial\mathcal{R}}{\partial i} + \frac{(1-e^2)^{1/2}}{ena^2}\frac{\partial\mathcal{R}}{\partial e}$$

$$= \frac{1}{a^2 n\sqrt{1-e^2}}\left[\frac{1-e^2}{e}\frac{\partial\mathcal{R}}{\partial e} - \frac{1}{\tan i}\frac{\partial\mathcal{R}}{\partial i}\right] \qquad (5)$$

(VI+V):

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{\sqrt{1-e^2}}{n}(\dot{\omega} + \dot{\Omega}\cos i) + \frac{2}{n^2 a}\frac{\partial\mathcal{R}}{\partial a} =$$

$$= \frac{1-e^2}{en^2a^2}\frac{\partial\mathcal{R}}{\partial e} + \frac{2}{n^2 a}\frac{\partial\mathcal{R}}{\partial a}.$$

Summary of Lagrange's planetary equations

→ In summary Lagrange's planetary equations (1-6) are

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -\frac{2}{n^2 a}\frac{\partial\mathcal{R}}{\partial\tau} \qquad (1) \qquad (I)$$

$$\frac{\mathrm{d}e}{\mathrm{d}t} = -\frac{1-e^2}{a^2en^2}\frac{\partial\mathcal{R}}{\partial\tau} - \frac{\sqrt{1-e^2}}{a^2en}\frac{\partial\mathcal{R}}{\partial\omega} \qquad (2) \qquad (II+I)$$

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{1}{a^2 n\sqrt{1-e^2}}\left[\frac{1}{\tan i}\frac{\partial\mathcal{R}}{\partial\omega} - \frac{1}{\sin i}\frac{\partial\mathcal{R}}{\partial\Omega}\right] \qquad (3) \qquad (III+II)$$

$$\frac{\mathrm{d}\Omega}{\mathrm{d}t} = \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial\mathcal{R}}{\partial i} \qquad (4) \qquad (V+IV)$$

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{1}{a^2 n\sqrt{1-e^2}}\left[\frac{1-e^2}{e}\frac{\partial\mathcal{R}}{\partial e} - \frac{1}{\tan i}\frac{\partial\mathcal{R}}{\partial i}\right] \qquad (5) \qquad (IV+V)$$

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{1-e^2}{ea^2n^2}\frac{\partial\mathcal{R}}{\partial e} + \frac{2}{n^2 a}\frac{\partial\mathcal{R}}{\partial a} \qquad (6') \qquad (VI+V)$$

→ The specific form of Lagrange's planetary equations depends on the choice of the elements.

→ Roy uses as orbital elements

a (semi-major axis)

e (eccentricity)

i (inclination)

Ω (longitude of the ascending node)

ω (argument of pericentre)

$\chi = -n\tau$ (mean anomaly at epoch, sometimes indicated with $\mathcal{M}_0$)

→ MD (see also Roy 7.7) use as orbital elements

a (semi-major axis)

e (eccentricity)

i (inclination)

Ω (longitude of the ascending node)

$\varpi = \Omega + \omega$ (longitude of pericentre)

$\epsilon = \varpi + \chi$ (mean longitude at epoch)

4.2.2 Expansion and classification of arguments of the disturbing function

[MD 6.3, 6.7, 6.9]

→ Lagrange's planetary equations can be studied analytically by writing the disturbing function $\mathcal{R}$ as a function of the orbital elements of the osculating ellipse and expanding the derivatives of $\mathcal{R}$ with respect to the elements.

→ Consider, for simplicity, the case of a three-body system. The motion of a body of mass m around a primary of mass m_0 is perturbed by the presence of a third body of mass m'. Viceversa, the motion of the body m' around the primary m_0 is perturbed by the presence of the body m.

→ The osculating elements are a, e, i, λ, ϖ, Ω for body m and a', e', i', λ', ϖ', Ω' for body m'. We use here, instead of τ and ω, the mean longitude λ and the longitude of pericentre ϖ, where $\lambda = n(t-\tau) + \varpi$ and $\varpi = \omega + \Omega$

→ Note that, in the unperturbed problem, the mean longitudes λ and λ' increase linearly with time, being proportional to nt and $n't$, while all the other elements are constant in the unperturbed problem. λ' and λ are the only rapidly varying variables.

→ Consider the motion of body m: it can be described using Lagrange's planetary equations. It can be shown (e.g. MD 6.3) that the perturbing function, as a function of the osculating elements, can be expanded in the form

$$\mathcal{R} = \sum_k S_k(a, a', e, e', i, i') \cos\phi_k,$$

where each ϕ_k is in the form

$$\phi = j_1\lambda' + j_2\lambda + j_3\varpi' + j_4\varpi + j_5\Omega' + j_6\Omega,$$

where j_i are positive and negative integers such that $\sum_{i=1}^{6} j_i = 0$.

→ We refer to the terms ϕ_k as the arguments of the expansion and to S_k as the coefficients of the expansion.

→ It is clear that, in the perturbed problem, all arguments in which λ and λ' do not appear correspond to slowly varying terms, or *secular terms* (i.e. long-period terms).

→ Among the terms whose arguments depend on λ and λ' we distinguish *resonant terms* and *short-period terms*. Note that in the unperturbed problem $j_1\lambda' + j_2\lambda = (j_1 n' + j_2 n)t + const.$

→ *Resonant terms* are those such that $j_1 n' + j_2 n \approx 0$, which happens when n and n' are commensurable (because the j_i are integers), i.e. when there is mean motion resonance. Recalling the relation $n \propto a^{-3/2}$, this condition for resonance can be expressed for semi-major axes as

$$\frac{a}{a'} \approx \left(\frac{|j_1|}{|j_2|}\right)^{2/3}$$

→ *Short-period terms* are all the other terms (those depending on λ and λ', but not resonant).

→ *Averaging principle:* the short-period terms average out to zero over the long period motion, so they can be neglected. The only important terms are the secular terms and the resonant terms.

→ Comparison with numerical integration support the idea that the averaging principle works. See Fig. MD 6.3 (FIG CM4.1 only secular terms, for a case far from resonance); Fig. MD 6.5 (FIG CM4.2 secular and resonant terms, for an asteroid close to 2:1 Jovian resonance)

4.2.3 The disturbing function for a system of N point masses*

[R05 7.10]

→ Let us write the equations of motion for a system of $N > 2$ bodies in the frame of reference in which one of the bodies (the reference body, with mass m_1) is in the origin (in this reference frame we call the position vector $\mathbf{r}$). Let us start from a reference system with origin in the centre of mass (in this reference frame we call the position vector $\mathbf{r}'$). For the i-th body we have:

$$m_i \ddot{\mathbf{r}}'_i = G \sum_{j=1,N}^{j \neq i} \frac{m_i m_j}{r_{ij}^3} \mathbf{r}_{ij},$$

where $\mathbf{r}_{ij} \equiv \mathbf{r}'_j - \mathbf{r}'_i$. So, for the reference body

$$\ddot{\mathbf{r}}'_1 = G \sum_{j=2,N} \frac{m_j}{r_{1j}^3} \mathbf{r}_{1j},$$

while for i-th body:

$$\ddot{\mathbf{r}}'_i = G \sum_{j=1,N} \frac{m_j}{r_{ij}^3} \mathbf{r}_{ij} \qquad (j \neq i)$$

*Optional for academic year 2013-2014

→ Subtracting we get

$$\ddot{\mathbf{r}}_{1i} = \ddot{\mathbf{r}}'_i - \ddot{\mathbf{r}}'_1 = G \sum_{j=2,N}^{j\neq i} \frac{m_j}{r_{ij}^3}\mathbf{r}_{ij} + G\frac{m_1}{r_{i1}^3}\mathbf{r}_{i1} - G \sum_{j=2,N}^{j\neq i} \frac{m_j}{r_{1j}^3}\mathbf{r}_{1j} - G\frac{m_i}{r_{1i}^3}\mathbf{r}_{1i},$$

→ Note that $\mathbf{r}_{ij} = -\mathbf{r}_{ji}$, we get

$$\ddot{\mathbf{r}}_{1i} = G \sum_{j=2,N}^{j\neq i} \frac{m_j}{\mathbf{r}_{ij}^3}\mathbf{r}_{ij} - G \sum_{j=2,N}^{j\neq i} \frac{m_j}{\mathbf{r}_{1j}^3}\mathbf{r}_{1j} - G\frac{m_i + m_1}{\mathbf{r}_{1i}^3}\mathbf{r}_{1i},$$

because

$$\frac{Gm_1}{r_{i1}^3}\mathbf{r}_{i1} = -\frac{Gm_1}{r_{1i}^3}\mathbf{r}_{1i}.$$

→ Finally, dropping the subscript 1, we get the equation of motion of mass m_i relative to the position of the reference mass m:

$$\ddot{\mathbf{r}}_i = -G(m+m_i)\frac{\mathbf{r}_i}{r_i^3} + G \sum_{j=2,N}^{j\neq i} m_j \left(\frac{\mathbf{r}_j - \mathbf{r}_i}{r_{ij}^3} - \frac{\mathbf{r}_j}{r_j^3}\right),$$

where $\mathbf{r}_{ij} = \mathbf{r}'_j - \mathbf{r}'_i = \mathbf{r}_j - \mathbf{r}_i$

→ In this equation the first term in the r.h.s. is the acceleration on body m_i due to the main body m. The other terms are (I: direct terms) acceleration on body m_i due to bodies m_j and (II: indirect terms) the negative of the acceleration on body m due to bodies m_j (which have effect on the motion of m_i, because we are taking as reference the position of the main body m).

→ In many cases the first term in the r.h.s. dominates. For instance, in the Solar System even Jupiter has $m_j/m \sim 0.001$. In the case of the Earth(m)-Moon(m_i)-Sun(m_j) system, the Earth-Moon distance ($\sim 3.84 \times 10^8\,\mathrm{m}$) is about 1/400 AU ($1\,\mathrm{AU} \sim 1.5 \times 10^{11}\,\mathrm{m}$), so the term $\sum_j$ is the sum of differences of almost equal numbers, so it is very small as compared to the first term (even though $M_\odot \sim 3 \times 10^5 M_{\rm Earth} \sim 3 \times 10^7 M_{\rm Moon}$).

→ The above equation can be written as

$$\ddot{\mathbf{r}}_i = \nabla_i \left(U_{0i} + \mathcal{R}_i\right),$$

where

$$U_{0i} \equiv \frac{G(m+m_i)}{r_i},$$

$$\mathcal{R}_i \equiv G \sum_{j=2,N}^{j\neq i} m_j \left(\frac{1}{r_{ij}} - \frac{\mathbf{r}_i \cdot \mathbf{r}_j}{r_j^3}\right),$$

and

$$\nabla_i = \frac{\partial}{\partial x_i}\hat{\imath} + \frac{\partial}{\partial y_i}\hat{\jmath} + \frac{\partial}{\partial z_i}\hat{k}.$$

→ Note that

$$r_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2},$$

$$\mathbf{r}_i \cdot \mathbf{r}_j = x_i x_j + y_i y_j + z_i z_j,$$

$$r_j = (\mathbf{r}_j \cdot \mathbf{r}_j)^{1/2} = \sqrt{x_j^2 + y_j^2 + z_j^2},$$

so

$$\nabla_i \left(\frac{1}{r_{ij}} \right) = \frac{\mathbf{r}_{ij}}{r_{ij}^3},$$

$$\nabla_i \left(-\frac{\mathbf{r}_i \cdot \mathbf{r}_j}{r_j^3} \right) = -\frac{\mathbf{r}_j}{r_j^3}.$$

→ Therefore

$$\nabla_i U_{0i} = -\frac{G(m + m_i)\mathbf{r}_{ij}}{r_{ij}^3},$$

$$\nabla_i \mathcal{R}_i = G \sum_{j=2,N}^{j \neq i} m_j \left(\frac{\mathbf{r}_{ij}}{r_{ij}^3} - \frac{\mathbf{r}_j}{r_j^3} \right),$$

→ $\mathcal{R}_i$ is the disturbing function for body m_i.

4.3 Special perturbations

[R05 8.1-8.2]

→ The method of special perturbations consists in numerically integrating the equations of motion of the N bodies in any of their possible forms. Also known as method of numerical integration of orbits.

→ In celestial mechanics the number of bodies is small, so computing the force is not expensive. Main limitation is the rounding-off error, which affects the long-term evolution of given initial conditions.

→ There are several implementations of the special perturbation method, depending on the formulation of the equations of motion and on the numerical integration algorithm. The choice depends on several factors: type of orbit, required accuracy, length of time span, available computing facilities.

4.3.1 Formulation of the equations of motion

[R05 8.3-8.5]

→ We recall here three of the main approaches: Cowell's method, Encke's method, integration of the perturbation equations (i.e. Lagrange's planetary equations).

→ *Cowell's method.* In this case the equations of motion are integrated in their usual form: Newton equations in Cartesian coordinates. So, for a system of N bodies, we have

$$m_i \ddot{\mathbf{r}}_i = -\frac{\partial V}{\partial \mathbf{r}_i}, \qquad i = 1, \ldots, N,$$

i.e.

$$m_i \ddot{x}_i = -\frac{\partial V}{\partial x_i}, \qquad m_i \ddot{y}_i = -\frac{\partial V}{\partial y_i}, ..., etc.,$$

where

$$V = -\frac{1}{2} G \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{m_i m_j}{r_{ij}},$$

with $j \neq i$, $r_{ij} = |\mathbf{r}_{ij}|$ and $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$. So

$$m_i \ddot{\mathbf{r}}_i = G \sum_{j=1,N}^{j \neq i} \frac{m_i m_j}{r_{ij}^3} \mathbf{r}_{ij},$$

or

$$\ddot{\mathbf{r}}_i = G \sum_{j=1,N}^{j \neq i} \frac{m_j}{r_{ij}^3} \mathbf{r}_{ij}.$$

So the Cartesian components are

$$\ddot{x}_i = G \sum_{j=1,N}^{j \neq i} \frac{m_j}{r_{ij}^3} (x_j - x_i), \qquad \text{etc.}$$

→ *Encke's method.* The equations of motion are written in the following form. Let us consider for instance the case of the Solar System. For a given planet of mass m, let us define $\mathbf{r}_{2\mathrm{b}}(t)$ as the planet-Sun separation vector in the two-body orbit of the planet due to the presence of the Sun only:

$$\ddot{\mathbf{r}}_{2\mathrm{b}} + \mu \frac{\mathbf{r}_{2\mathrm{b}}}{r_{2\mathrm{b}}^3} = 0,$$

where $\mu = G(M_\odot + m)$. The actual orbit is

$$\ddot{\mathbf{r}} + \mu \frac{\mathbf{r}}{r^3} = \mathbf{F},$$

where $\mathbf{F}$ is the force due to all the other bodies (but the Sun). Subtracting, we have

$$\ddot{\boldsymbol{\rho}} = \mathbf{F} + \mu \left(\frac{\mathbf{r}_{2\mathrm{b}}}{r_{2\mathrm{b}}^3} - \frac{\mathbf{r}}{r^3} \right)$$

where $\boldsymbol{\rho} \equiv \mathbf{r} - \mathbf{r}_{2\mathrm{b}}$. The term in parenthesis is algebraically manipulated to avoid the necessity of using too many significant figures, due to subtraction of very similar numbers. $\mathbf{r}_{2\mathrm{b}}(t)$ is known analytically and the integration variable is the small difference $\boldsymbol{\rho}$, which varies more slowly than $\mathbf{r}$. This is exploited to use longer timesteps in the integration.

→ *Integration of perturbation equations.* Lagrange's planetary equations are integrated numerically. Advantages: longer timestep than equations of motion integration; orbital elements in input/output; no need to include explicitly central body (perturbation method). Disadvantages: complicated functions in the right-hand side; breakdown of the equation in specific configurations (denominators to zero); necessity to solve Kepler's equation

$$\mathcal{M} = \xi - e \sin \xi$$

which allows to convert the mean anomaly into the eccentric anomaly ξ, needed in order to obtain the position of the body, using distance from the focus r and true anomaly f, given by

$$r = a(1 - e \cos \xi)$$

and

$$\cos f = \frac{\cos \xi - e}{1 - e \cos \xi}.$$

4.3.2 Numerical integration algorithms

[P92 chapter 16; B07]

→ In all approaches we have to integrate a system of ordinary differential equations (ODEs). These might be first-order (Lagrange's planetary equations) or second order (Encke's or Cowell's equations). But a second-order system can be always recast in the form of a first-order system. Note that in some cases it is more convenient to integrate directly second order: for instance Stoermer rule can be applied to the full N-body equations because the right-hand side does not depend on the first derivatives (force is conservative, i.e. it depends only on position; see P92 16.5). However, here we consider only the case of first-order systems.

→ A second-order ODE (or system of ODEs) can be transformed into a first-order ODE (or system of ODEs) as follows:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = F(t, y, y'),$$

where $y' \equiv \mathrm{d}y/\mathrm{d}t$.

$$v(t) \equiv \frac{\mathrm{d}y}{\mathrm{d}t}$$

so we get the system

$$\frac{\mathrm{d}y}{\mathrm{d}t} = v$$

$$\frac{\mathrm{d}v}{\mathrm{d}t} = F(t, y, v)$$

with $v = v(t)$ and $y = y(t)$. Writing $\mathbf{w} = (y, v)$, the above system is clearly in the form

$$\frac{\mathrm{d}\mathbf{w}}{\mathrm{d}t} = \mathbf{f}(t, \mathbf{w}),$$

where $\mathbf{w} = \mathbf{w}(t)$.

→ When the problem is reduced to first-order system of ODE, we must just find a method to solve an equation in the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t, y),$$

where $y = y(t)$. We recall that in the N-body case t is time and y is either a phase-space coordinate (Newton's equations of motion) or an orbital element (Lagrange's planetary equations). In all cases our problem is an initial value problem, so the initial values t_0 (i.e. initial time) and $y_0 = y(t_0)$ (initial coordinates or elements) are given.

→ Numerical methods to integrate a system of ODEs: several choices are possible. Well known algorithms are Runge-Kutta, Bulirsch-Stoer, symplectic integrators etc. Typically in celestial mechanics high accuracy is required. This is due to a combination of the chaotic nature of the orbits and the necessity of integrating over long time spans: if the integration is not accurate enough, relatively small integration errors can lead to completely wrong orbits over long timescales.

→ Here we present only the Bulirsch-Stoer method, which is a robust method, often used in applications of celestial mechanics.

4.3.3 Bulirsch-Stoer algorithm

[P92 16.1-16.4]

→ Bulirsch-Stoer algorithm is a method to integrate systems of ODE based on Euler method, Midpoint method and Modified midpoint method. Before describing the Bulirsch-Stoer algorithm, we briefly describe these simpler methods. We recall that we want to solve a first-order ODE in the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t, y),$$

where $y = y(t)$, with initial conditions $y = y_0$ for $t = t_0$. In general, given t_i and $t_{i+1} = t_i + H$ (where H is the integration step), for known $y_i = y(t_i)$ we want to estimate $y(t_{i+1})$ numerically: the approximated result is called $y_{i+1} \approx y(t_{i+1})$. Such integration steps are repeated to go from the initial value t_0 to the final value t_{final} of the independent variable (i.e. time, in the N-body case).

→ *Euler method.*

$$y_{i+1} = y_i + Hf(t_i, y_i),$$

where H is increment (or step) and $f = y'$ is evaluated at (t_i, y_i). y_{i+1} is the estimated value of $y(t_{i+1})$ where $t_{i+1} = t_i + H$.

→ *Midpoint method.* The step H is divided in two steps of length $H/2$ and the slope is evaluated at $t + H/2$.

$$y_{i+1} = y_i + Hf(t_{i+1/2}, y_{i+1/2}),$$

where

$$y_{i+1/2} \equiv y_i + \frac{H}{2} f(t_i, y_i),$$

and $t_{i+1/2} \equiv t_i + (H/2)$.

→ *Modified midpoint method.* The step H is divided in n steps of length $h = H/n$. Now $y_i = y(t_i)$ and $y_{i+1} \approx y(t+H)$. The algorithm reads as follows:

$$z_0 = y_i = y(t_i)$$

$$z_1 = z_0 + hf_0 \approx y(t_i + h),$$

which is an estimate of $y(t_i + h)$ using Euler's method. Here we have introduced the following notation: $f_j \equiv f(t_i + jh, z_j)$, so, $f_0 = f(t_i, z_0)$.

$$z_2 = z_0 + 2hf_1 \approx y(t_i + 2h),$$

which is an estimate of $y(t_i + 2h)$ using the midpoint method. In general, for the j-th sub-step we have

$$z_j = z_{j-2} + 2hf_{j-1} \approx y(t_i + jh),$$

which is an estimate of $y(t_i + jh)$ using the midpoint method. Finally we define y_{i+1} by averaging between z_n and the average between z_{n+1} and z_{n-1}:

$$y_{i+1} = \frac{1}{2}\left(\frac{z_{n+1} + z_{n-1}}{2} + z_n\right),$$

i.e.

$$y_{i+1} = \frac{1}{4}\left(z_{n-1} + 2z_n + z_{n+1}\right),$$

$$= \frac{1}{4}\left(z_{n-1} + 2z_n + z_{n-1} + 2hf_n\right),$$

i.e.

$$y_{i+1} = \frac{1}{2}\left[z_n + z_{n-1} + hf_n\right].$$

→ *Bulirsch-Stoer method.* With this method each step goes from t to $t + H$, via several (n) modified-midpoint method sub-steps with $h = H/n$, which are extrapolated to $h \to 0$.

→ n is not fixed, but for each step we try first with $n = 2$, and then increase n iteratively up to a value which is estimated to be sufficient (i.e., such that the error is small enough).

→ n is not increased by one each time, but through a specific sequence. One of the optimal choices is

$$n = 2, 4, 6, 8, 10, ... \qquad \text{(i.e. } n_k = 2k\text{)},$$

where k is the index that represents the iteration step.

→ For given k, so for given n_k (and then for given $h_k = H/n_k$), the modified midpoint method gives us an estimate $y_{i+1}(h_k)$, depending on h_k. For each k, i.e. each n_k in the sequence, via polynomial extrapolation, we compute

$$y_{i+1,k} = \lim_{h\to 0} g_k(h),$$

where $g_k(h)$ is a polynomial function interpolating the k points $[h_k, y_{i+1}(h_k)]$. This method is known as Richardson extrapolation.

→ The extrapolation can be performed as follows. Given k estimates of $y_{i+1,k}$, corresponding to k different values of n, we define an interpolating function, a polynomial of order $k-1$

$$g_k(h) = a_0 + a_1 h + a_2 h^2 + ... + a_{k-1} h^{k-1}.$$

There is only an interpolating polynomial of order $k-1$ (obtained, for instance, with Lagrange formula or other interpolating algorithm). So we can compute the coefficients. Then we can compute the extrapolation to $h = 0$, which is simply $y_{i+1,k} \equiv g_k(0) = a_0$

→ We go on for increasing k. We stop for $k = k'$ when we meet a convergence criterion. For instance

$$\frac{|y_{i+1,k'} - y_{i+1,k'-1}|}{|y_{i+1,k'}|} < \epsilon,$$

where ϵ is a (small) dimensionless number, which is the accuracy (e.g. $\epsilon \sim 10^{-13}$).

→ Finally

$$y_{i+1} = y_{i+1,k'} \equiv g_{k'}(0).$$

See plot B07 Fig. 3.4 (FIG CM4.3) and P92 Fig. 16.4.1 (FIG CM4.4).

Bibliography

→ Bodenheimer P. et al. 2007 “Numerical Methods in Astrophysics: An Introduction”, Taylor & Francis (B07)

→ Murray C.D., Dermott S.F., 1999, “Solar system dynamics”, Cambridge University Press, Cambridge (MD)

→ Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, “Numerical Recipes in Fortran, 2nd edn.”, Cambridge University Press, Cambridge (P92)

→ Roy A.E. 2005, “Orbital motion”, 2005 Taylor & Francis (R05)